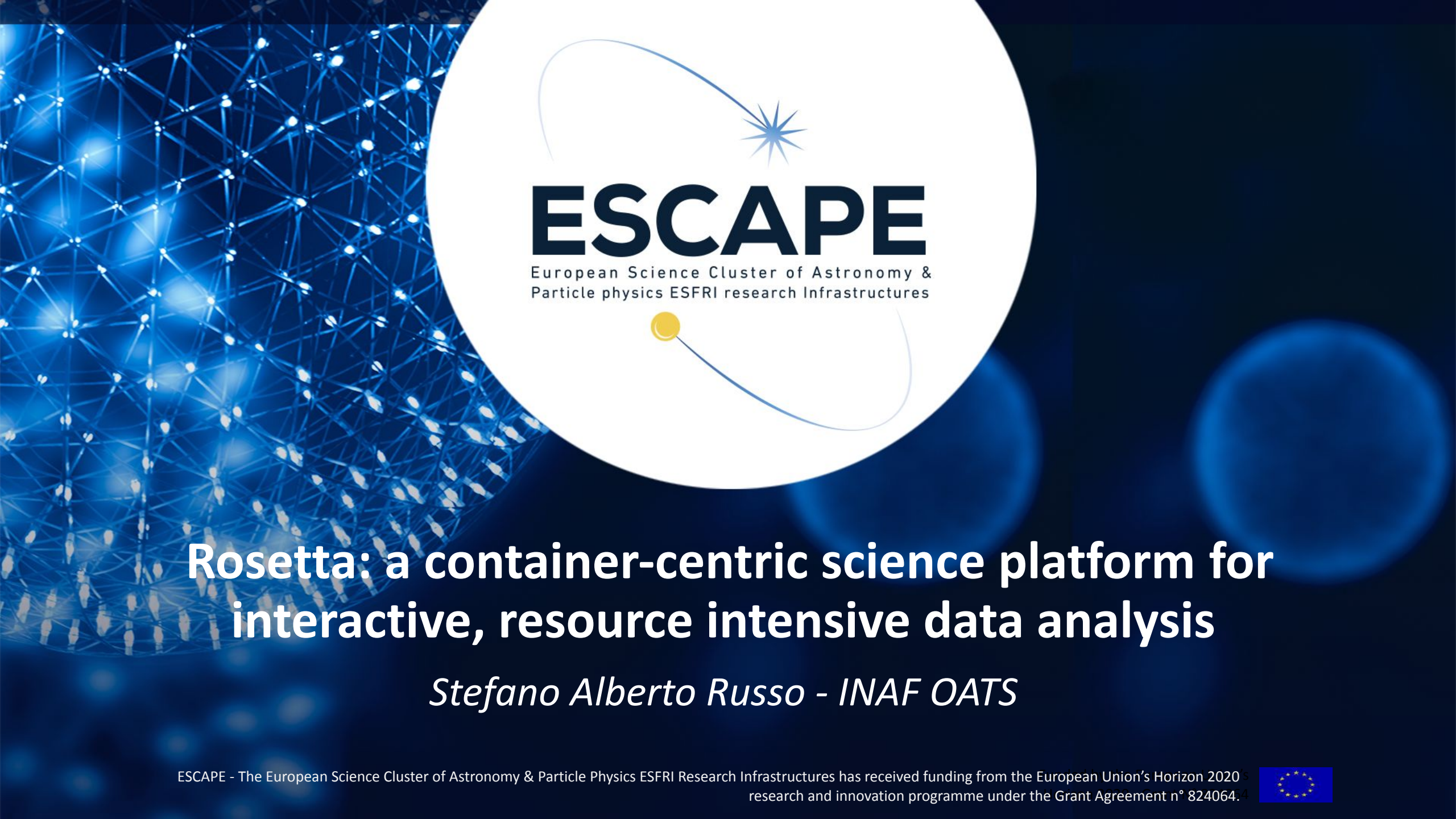


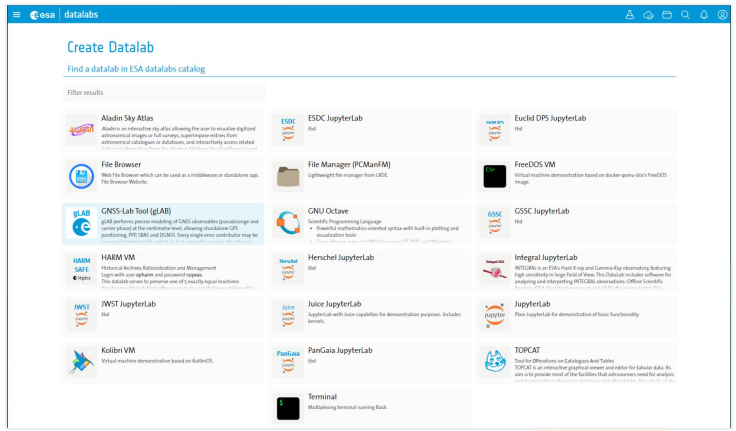
Rosetta: a container-centric science platform for interactive, resource intensive data analysis

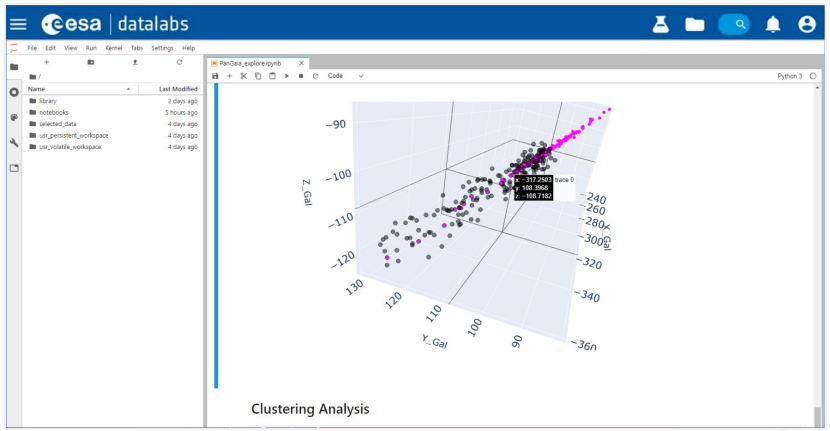
Stefano Alberto Russo - INAF OATS

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

What is a science platform?

"A science platform is an environment designed to offer users a smoother experience when interacting with computing and storage resources"





Credits: [ESA Data Labs]

Funded by the European Union's Horizon 2020 - Grant N° 824064

What is a science platform?

"A science platform is an environment designed to offer users a smoother experience when interacting with computing and storage resources"

- We usually assume a web-based environment built on top of Jupyter notebooks or similar software

- Other approaches or definitions may be also feasible (e.g. full interactive desktop access).

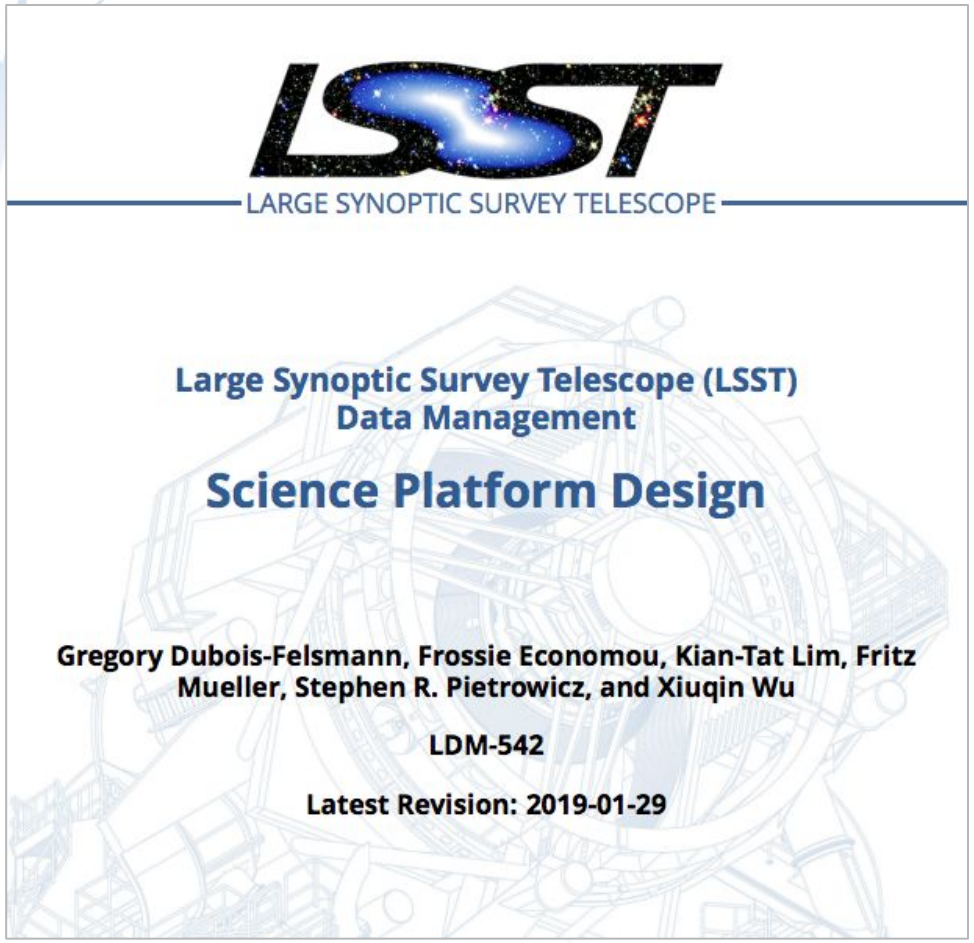
Credits: ESA Data Labs

Funded by the European Union's Horizon 2020 - Grant N° 824064

Example definition

"a web Portal, designed to provide essential data access and visualization services through a simple-to-use website, a Notebook environment, that will provide a Jupyter Notebook-like interface, based on JupyterLab, enabling next-to-the-data analysis”

Juric et al. 2017



Funded by the European Union’s Horizon 2020 - Grant N° 824064

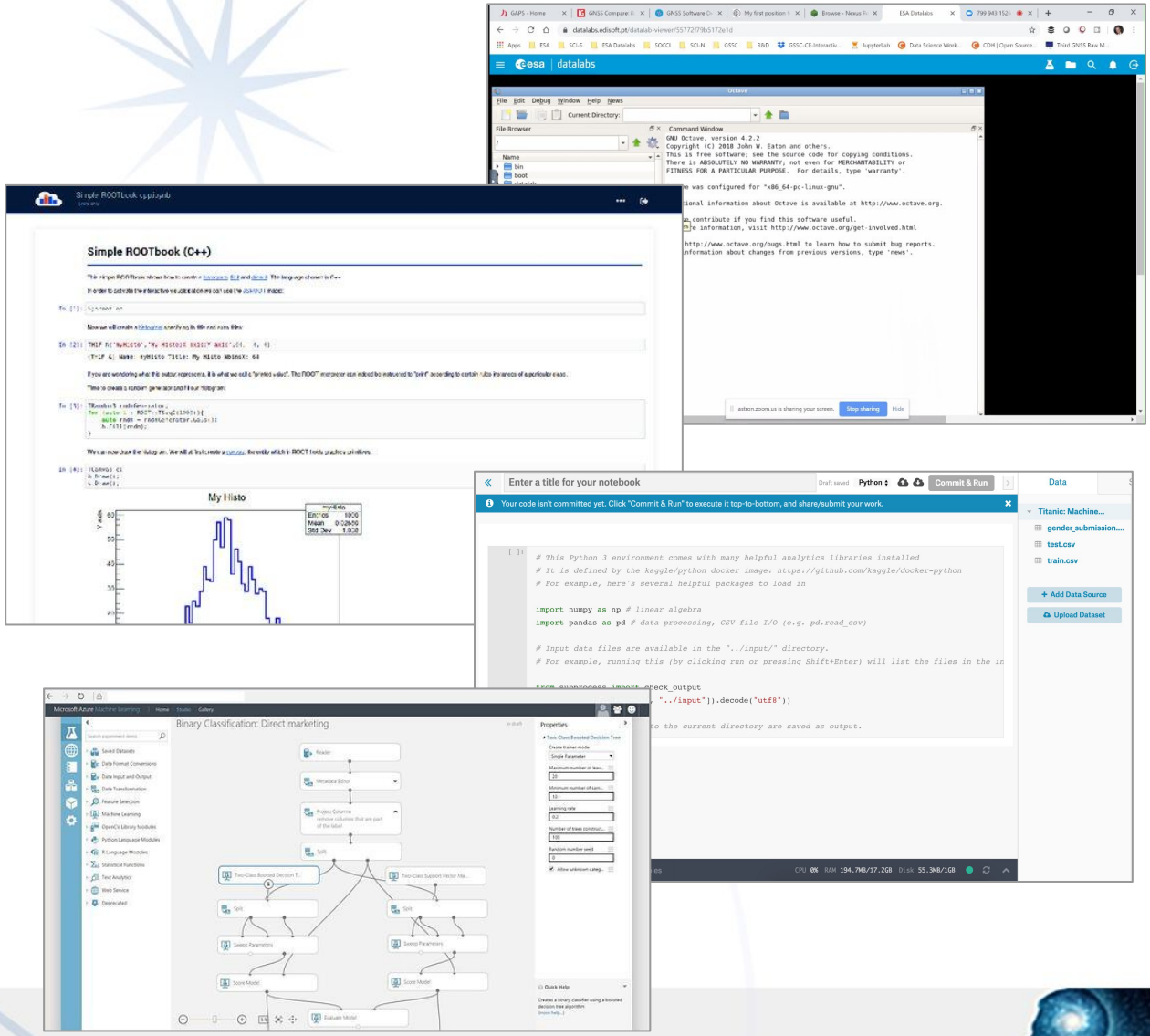
Some examples

Public Research/ Academia:

- CERN SWAN
- ESA Datalabs
- LSST Science Platform
- ...

Private Research / Industry:

- Google Colabs
- Kaggle Kernels
- Azure Machine Learning Studio
- ...



Funded by the European Union's Horizon 2020 - Grant N° 824064

Limitations

There are some strong limitations with the current science platforms:

1. A user cannot run in an environment not supported by the platform
   → from a different Python version to a different Linux distribution.
   → this is a reproducibility issue!!

2. A user is tight to a specific interface, with little options for other ones
   → a web-based notebook (Jupyter or similar) interface makes it impossible to run native GUI applications (common in Astrophysics).

3. Poor HPC support, hard to integrate
   → the focus is usually more on making the data accessible rather than to run resource-intensive analysis

Funded by the European Union's Horizon 2020 - Grant N° 824064

Introducing Rosetta

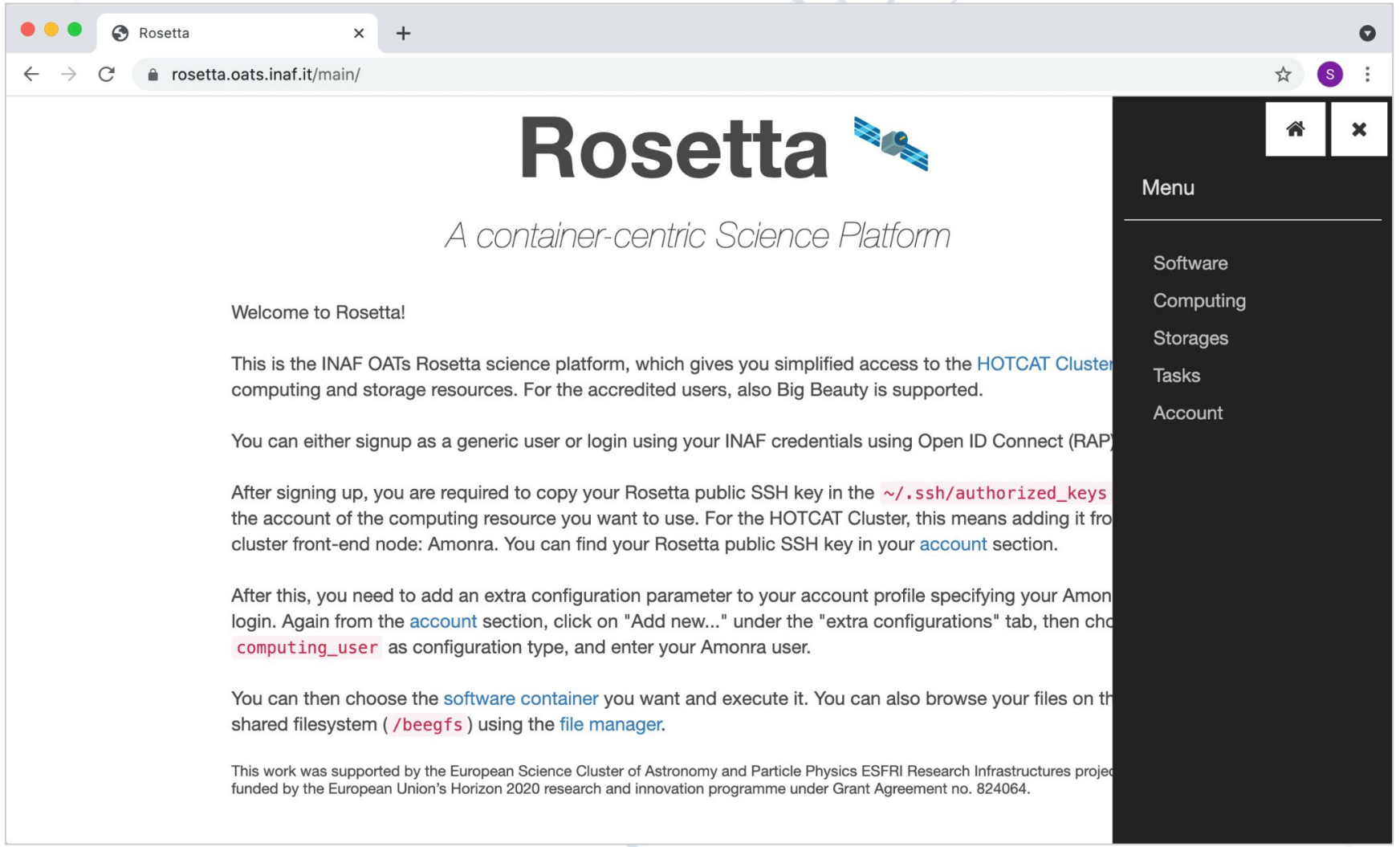
- A web-based science platform supporting both pre-defined and custom analysis environments
- Executes user tasks in software containers on:
  - local CPUs;
  - standalone computing resources
  - HPC clusters and Cloud systems
- Microservice-oriented approach: each task runs in its own container and expose its interface directly to the user:

  → *Users can choose from different software and interfaces*

  → *Users can add their own containers*

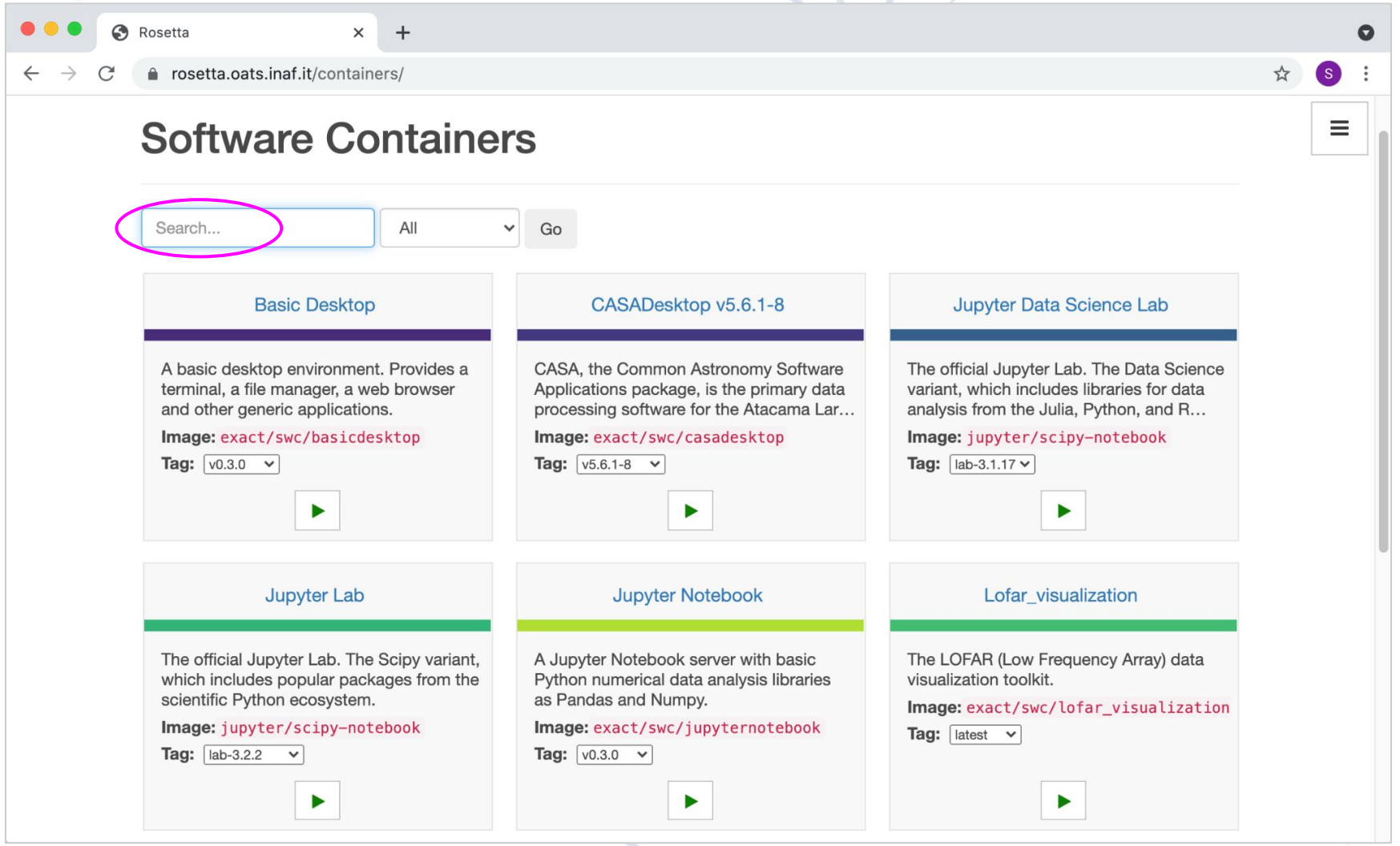
Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta main page



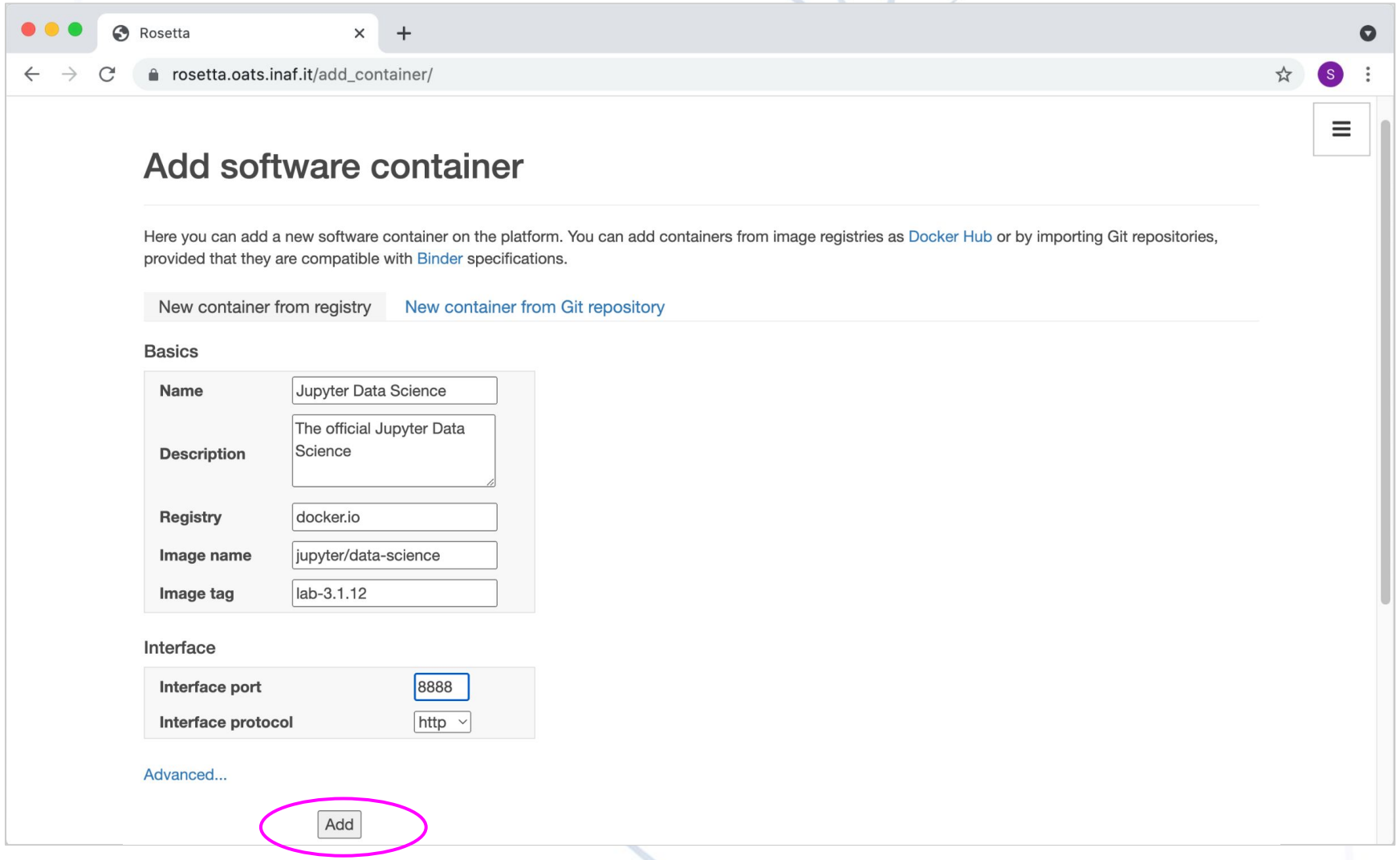
Funded by the European Union's Horizon 2020 - Grant N° 824064

The software “catalogue”



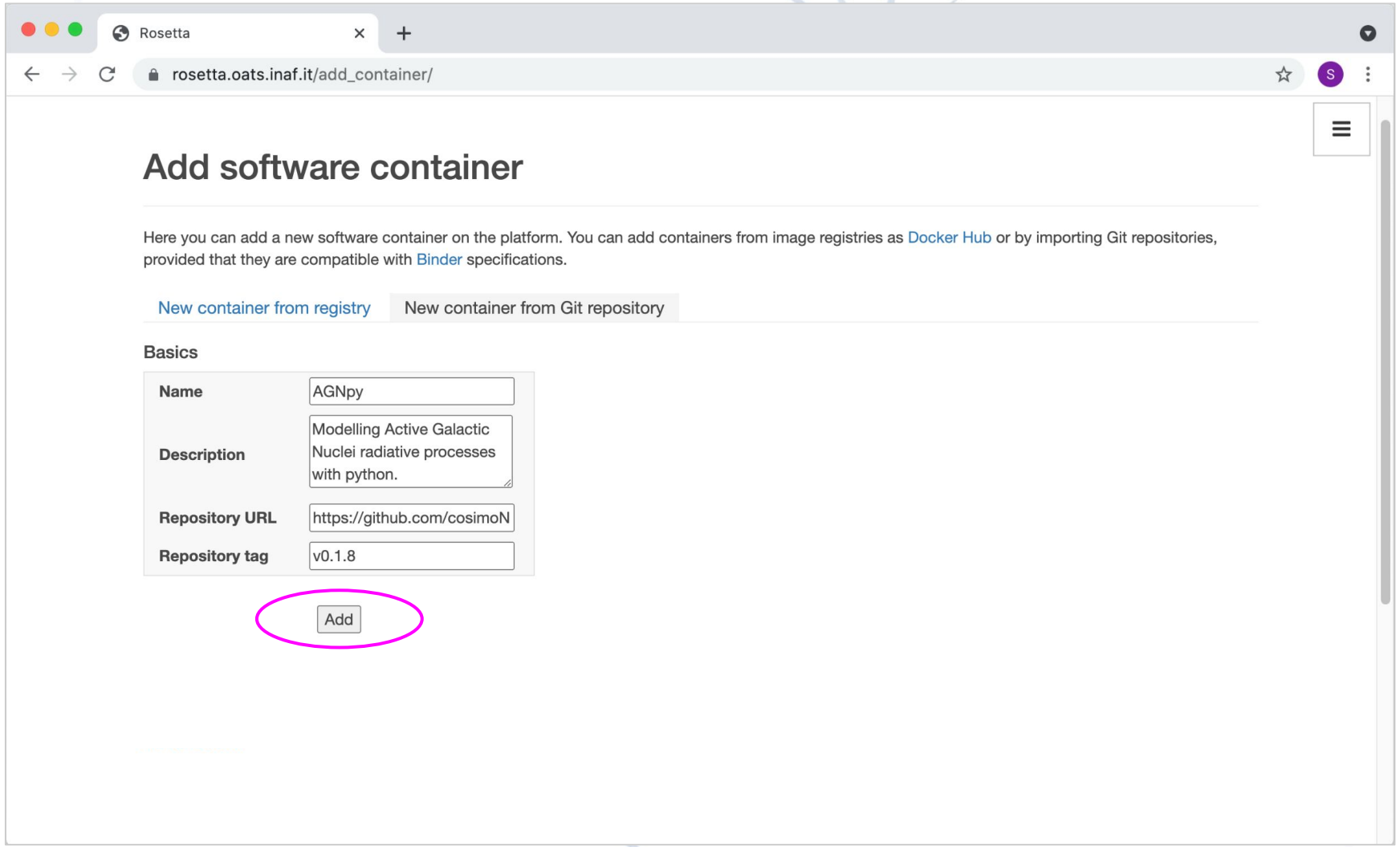
Funded by the European Union’s
Horizon 2020 - Grant N° 824064

Adding new software



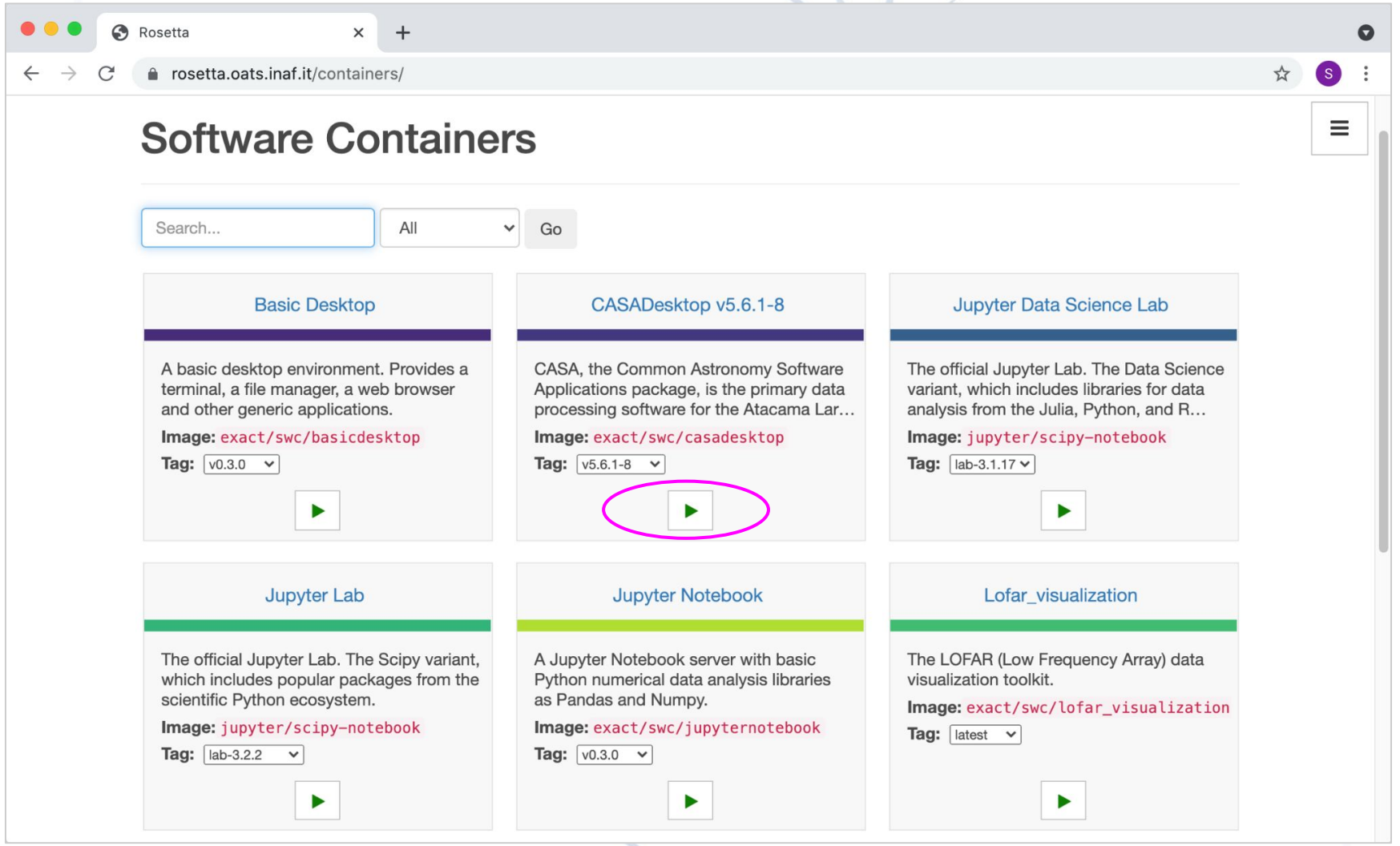
Funded by the European Union's Horizon 2020 - Grant N° 824064

Adding new software



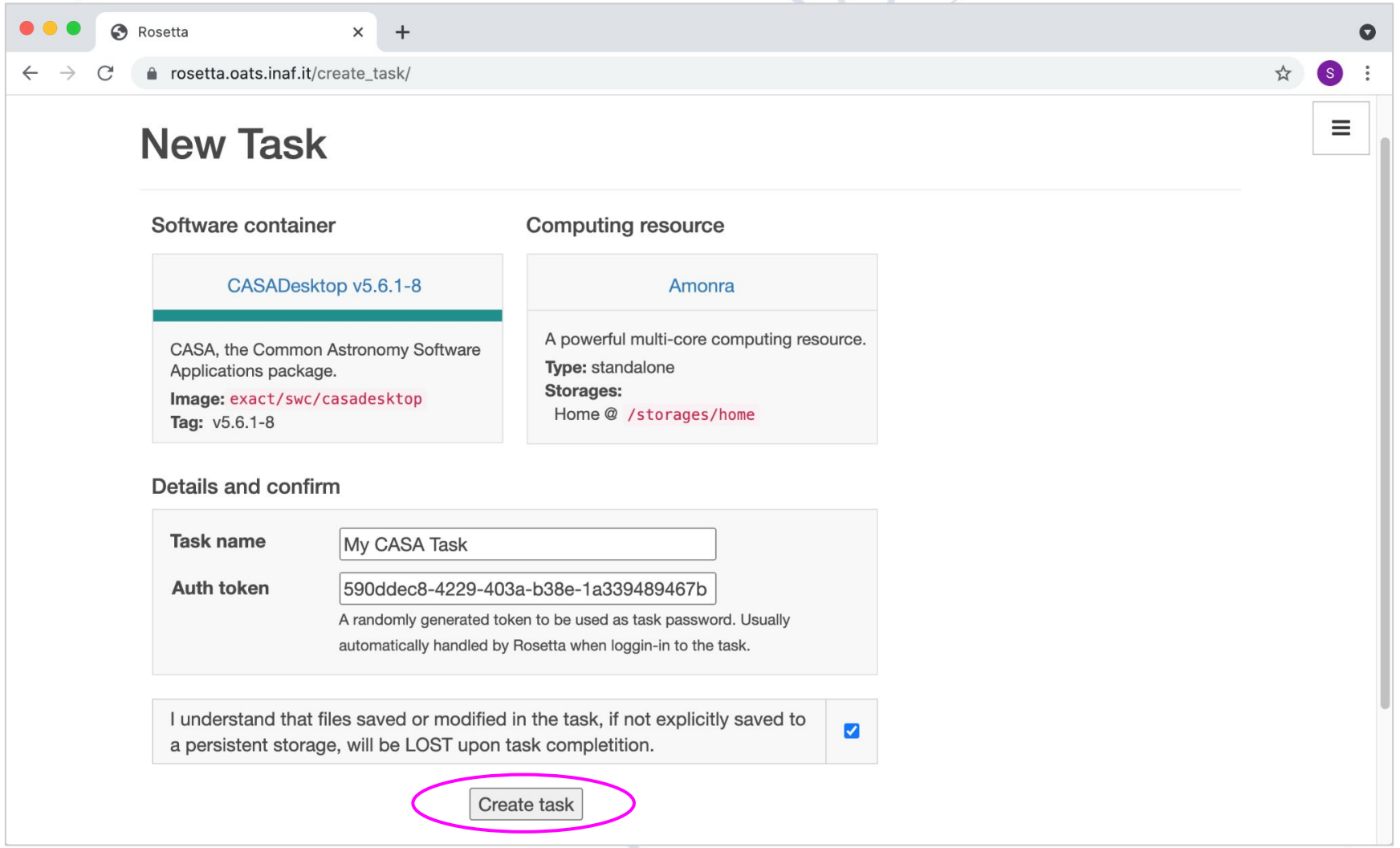
Funded by the European Union's Horizon 2020 - Grant N° 824064

Creating a new task



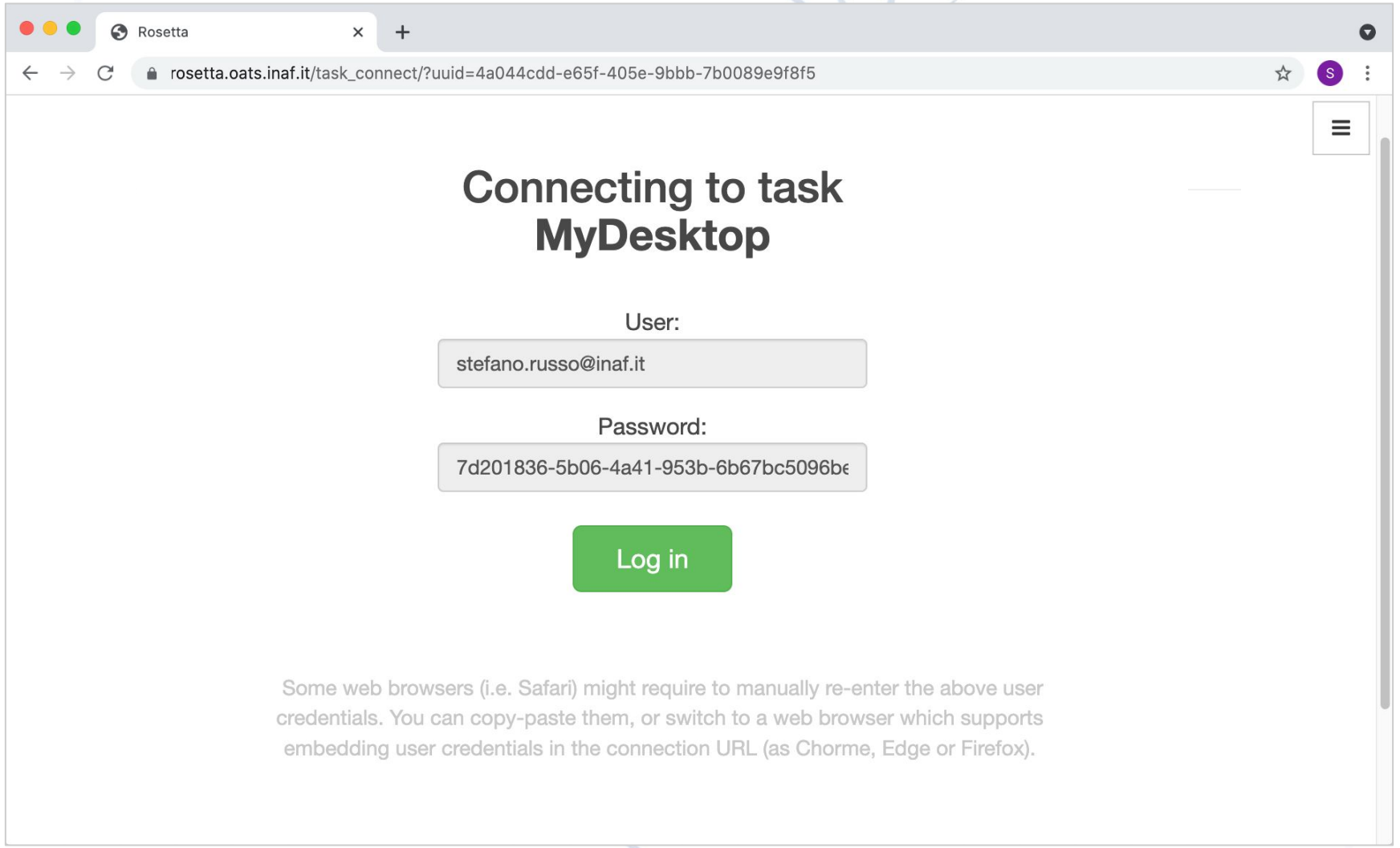
Funded by the European Union's Horizon 2020 - Grant N° 824064

Creating a new task



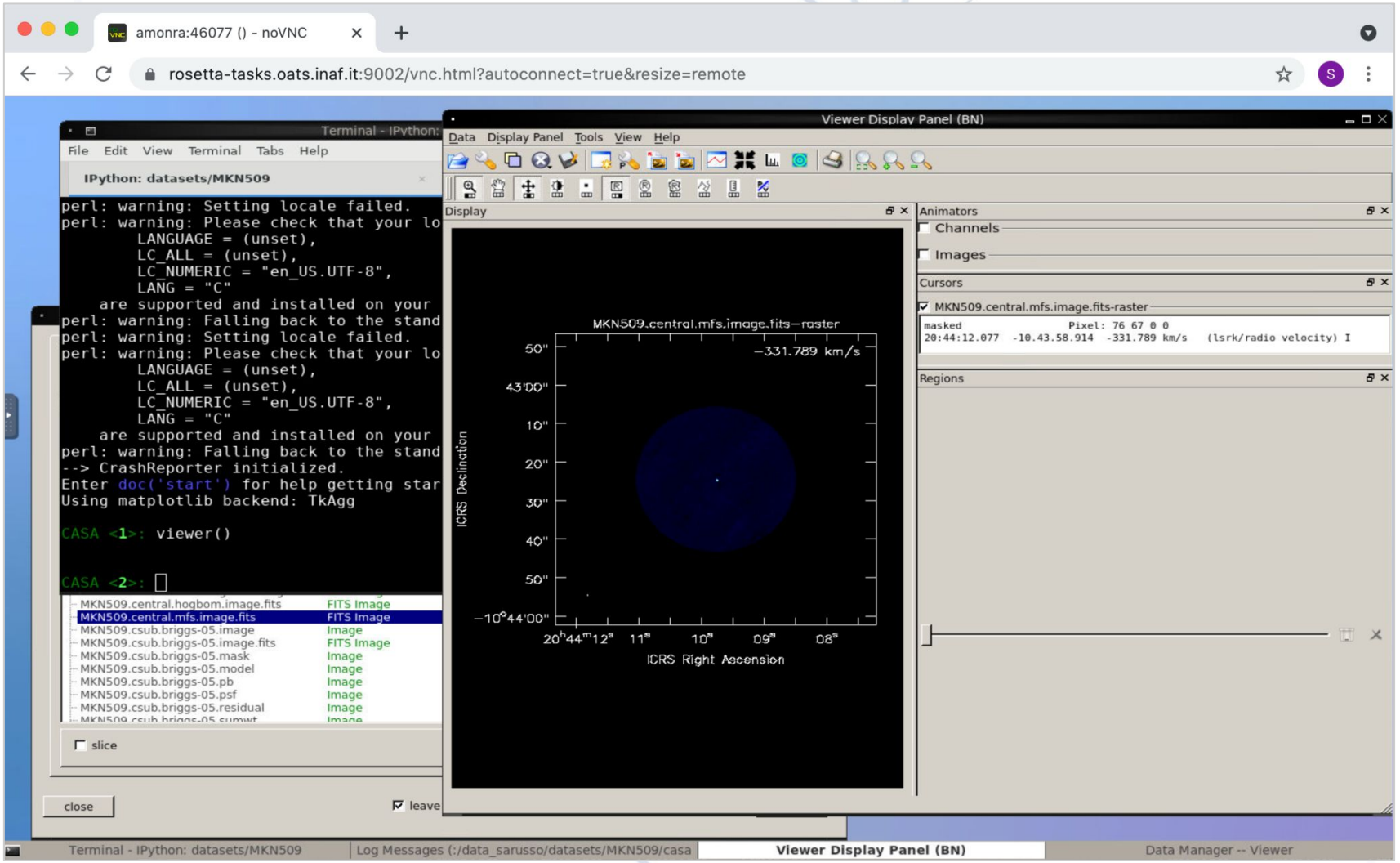
Funded by the European Union's Horizon 2020 - Grant N° 824064

Creating a new task



Funded by the European Union's Horizon 2020 - Grant N° 824064

A remote desktop task



Funded by the European Union's Horizon 2020 - Grant N° 824064

A Jupyter Notebook task

Test Notebook

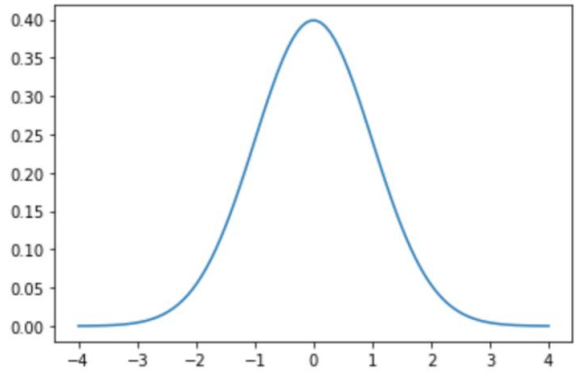
rosetta-tasks.oats.inaf.it:9001/notebooks/Test%20Notebook.ipynb

jupyter **Test Notebook** Last Checkpoint: 4 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Help Trusted Python 3

Run Code

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
mu = 0
std = 1
x = np.linspace(start=-4, stop=4, num=100)
y = stats.norm.pdf(x, mu, std)
plt.plot(x, y)
plt.show()
```

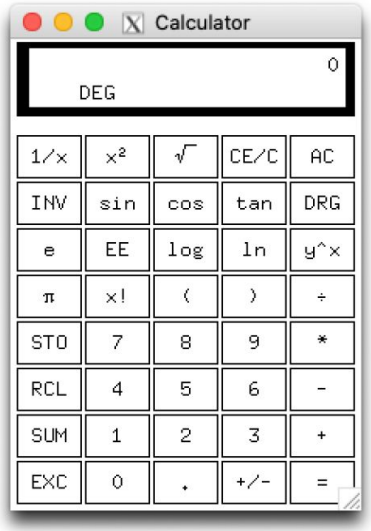


Funded by the European Union's Horizon 2020 - Grant N° 824064

A SSH server task

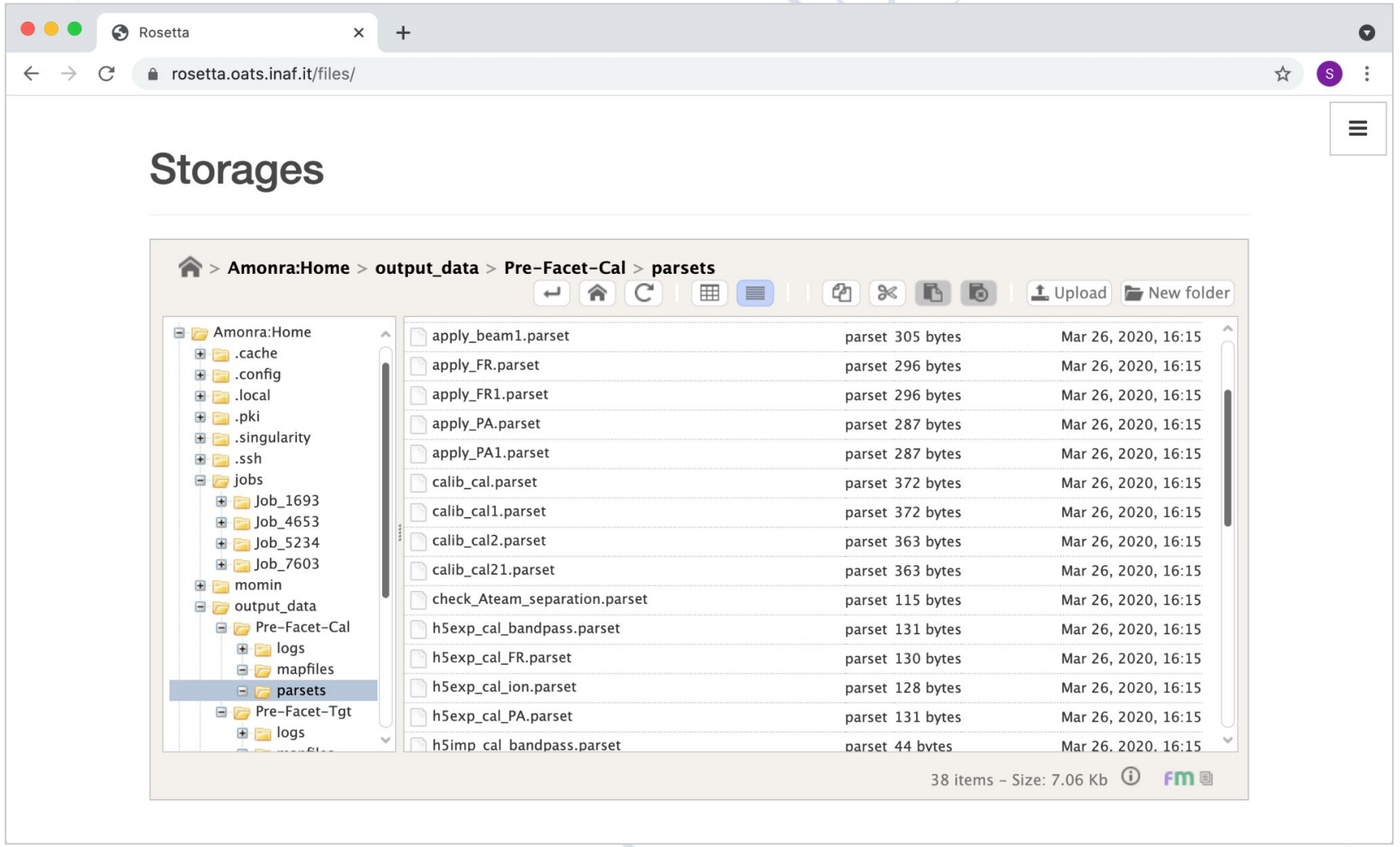
```
ste — ssh -X -p7000 metauser@rosetta-tasks.oats.inaf.it — 116×36
ste@Stes-MacBookAir:~ $ ssh -X -p7000 metauser@rosetta-tasks.oats.inaf.it
metauser@rosetta-tasks.oats.inaf.it's password:
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

metauser@task-0d378dfd:~$ pwd
/home/metauser
metauser@task-0d378dfd:~$ ls -alh
total 44K
drwxr-xr-x 6 metauser metauser 4.0K Nov 17 23:55 .
drwxrwxrwx 1 root     root     4.0K Nov 17 23:53 ..
-rw------- 1 metauser metauser   50 Nov 17 23:55 .Xauthority
-rw------- 1 metauser metauser  123 Nov 17 23:55 .bash_history
-rw-r--r-- 1 metauser metauser  220 Apr  4  2018 .bash_logout
-rw-r--r-- 1 metauser metauser 3.8K Nov 17 23:54 .bashrc
drwx------ 2 metauser metauser 4.0K Nov 17 23:53 .cache
-rw-r--r-- 1 metauser metauser    0 Nov 17 23:53 .initialized
drwxrwxr-x 3 metauser metauser 4.0K Nov 17 23:54 .local
drwxr-xr-x 2 metauser metauser 4.0K Nov  4 19:58 .logs
-rw-r--r-- 1 metauser metauser  807 Apr  4  2018 .profile
drwxr-xr-x 2 metauser metauser 4.0K Nov 17 23:53 custom_ssh
metauser@task-0d378dfd:~$ xcalc
```



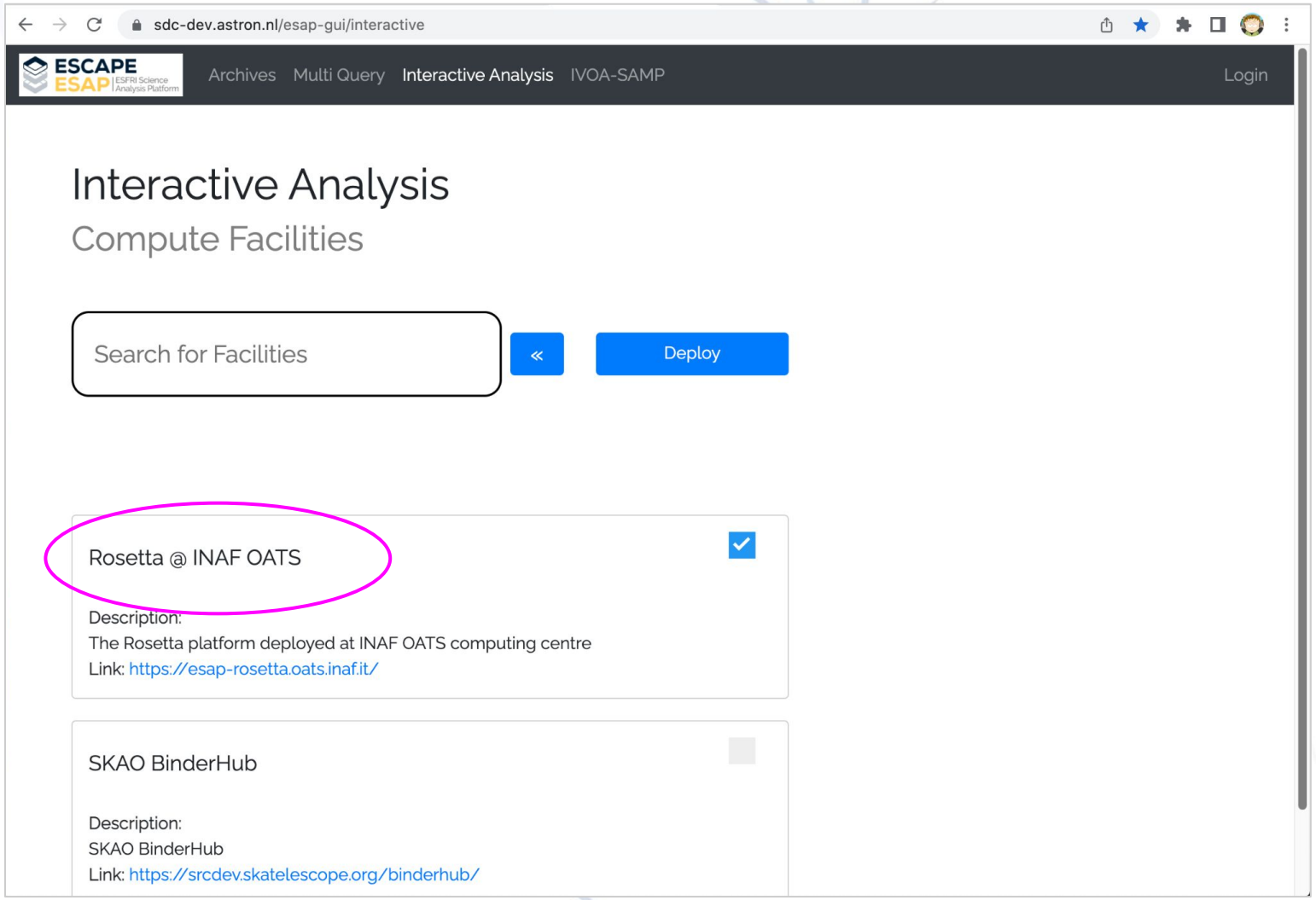
Funded by the European Union's Horizon 2020 - Grant N° 824064

The file manager



Funded by the European Union's Horizon 2020 - Grant N° 824064

ESAP integration



Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta architecture

Rosetta architecture is based on two main ingredients:

1) Microservices

2) Software containers

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Rosetta architecture: microservices

Microservices are independent and self-contained units that perform a given and well-defined task, using a well defined *interface*.

→ From just summing two numbers to running a neural network.

Microservices are completely decoupled from the underlying infrastructure and from each other. Encapsulation is maximum.

In Rosetta, each user task is framed as a microservice

Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta architecture: microservices

Examples of interfaces:

- REST APIs
- HTTP
- SSH
- RPC
- etc.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Rosetta architecture: containers

Software containers are lightweight, standalone, executable packages of software that includes everything needed to run an application:

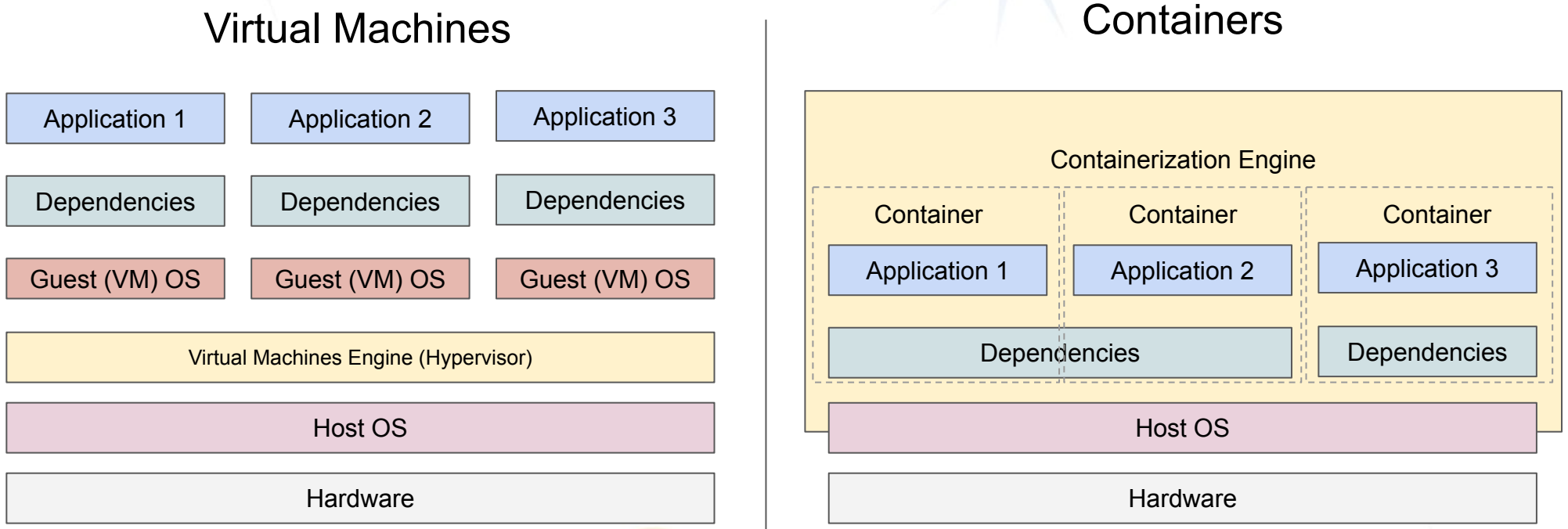
→ code, runtime, system tools, system libraries and settings.

They are a solution to the problem of how to get software to run reliably when moved from one computing environment to another.

Rosetta uses them to wrap the user task microservices

Funded by the European Union's Horizon 2020 - Grant N° 824064

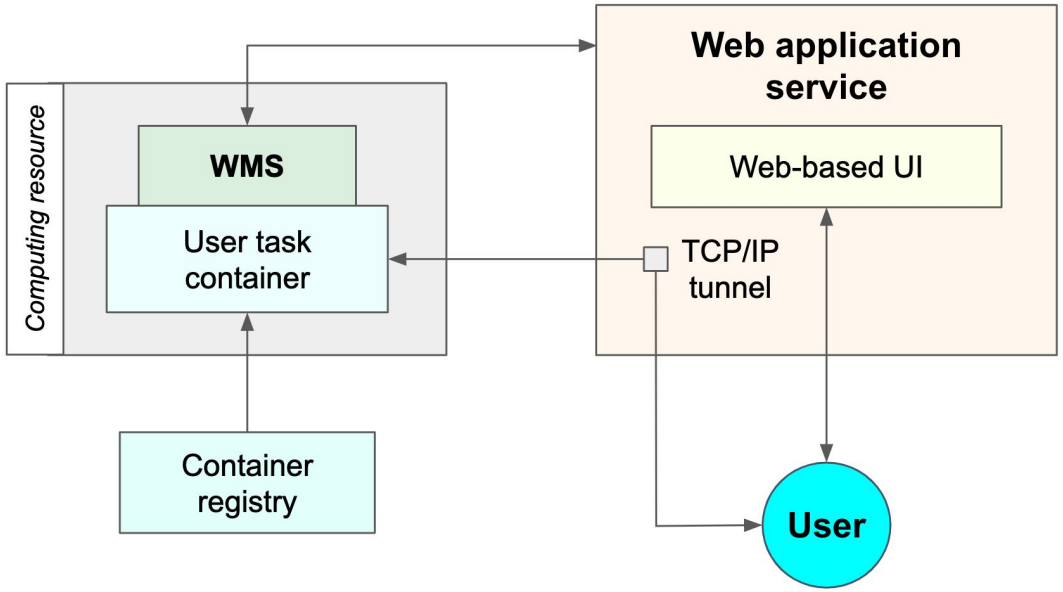
Rosetta architecture: containers

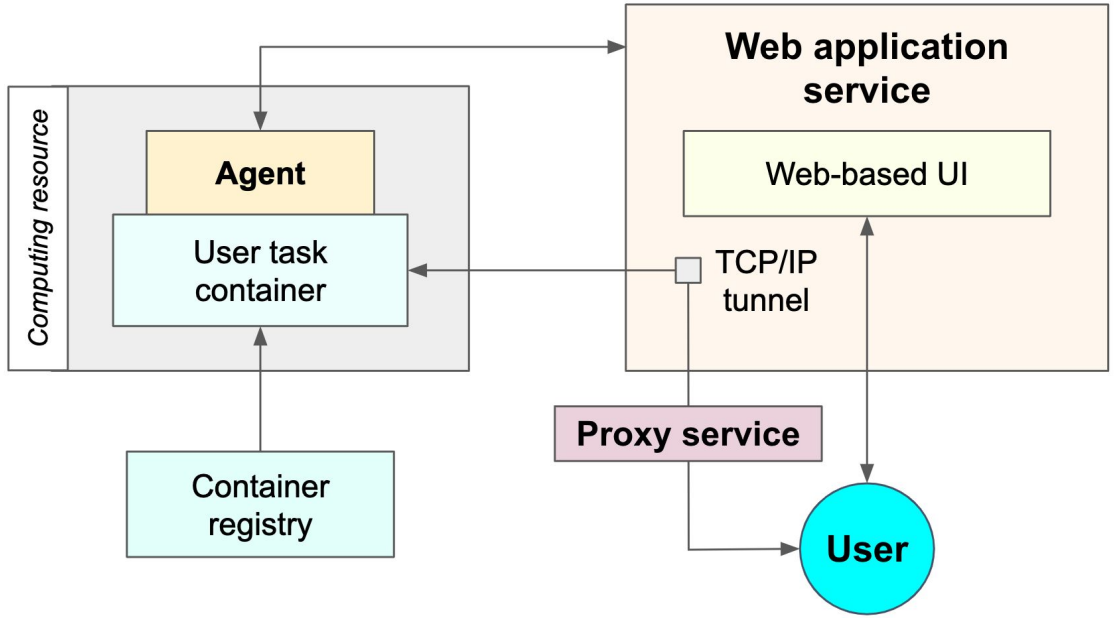


Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta architecture: task orchestration

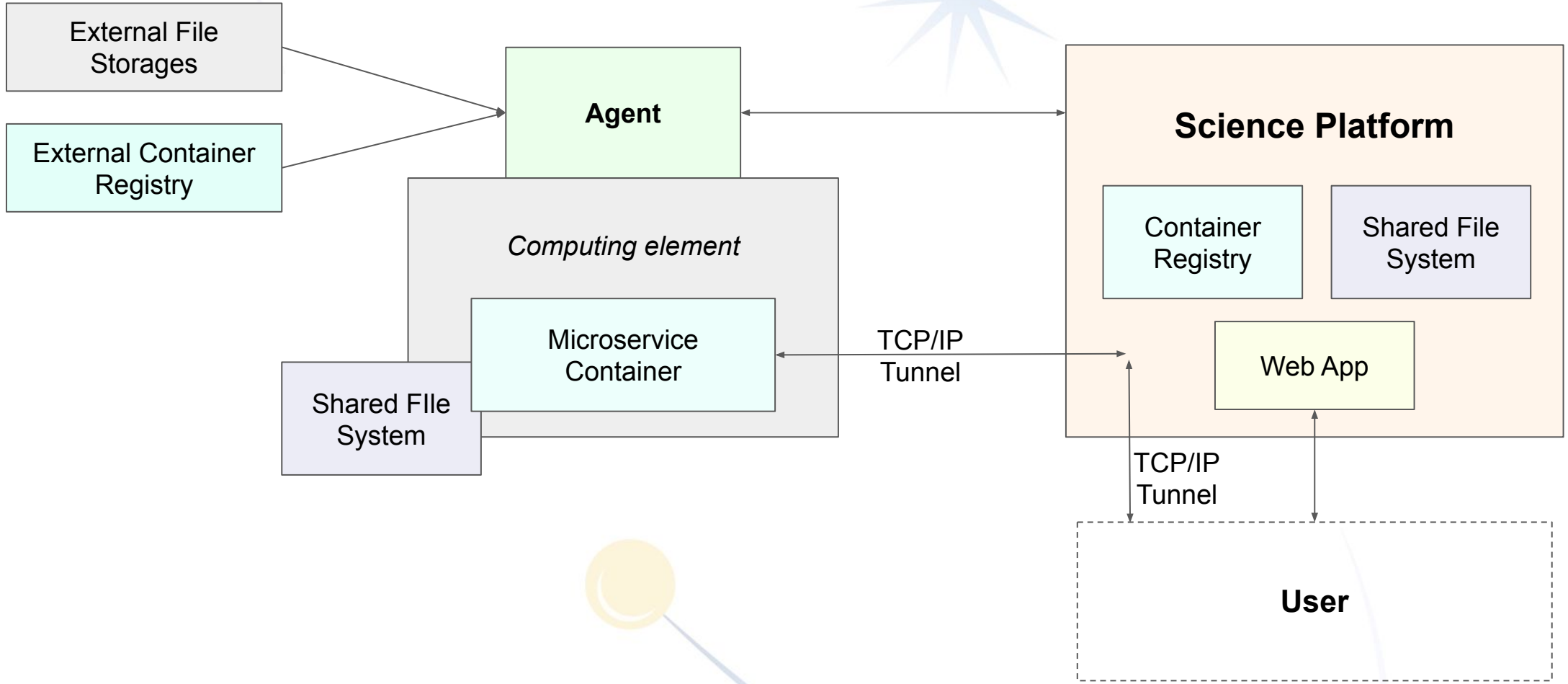
Task orchestration include sending the task for execution and setting up the all the tunneling required to reach its interfaces. It can rely on the WMS or use an *agent*.





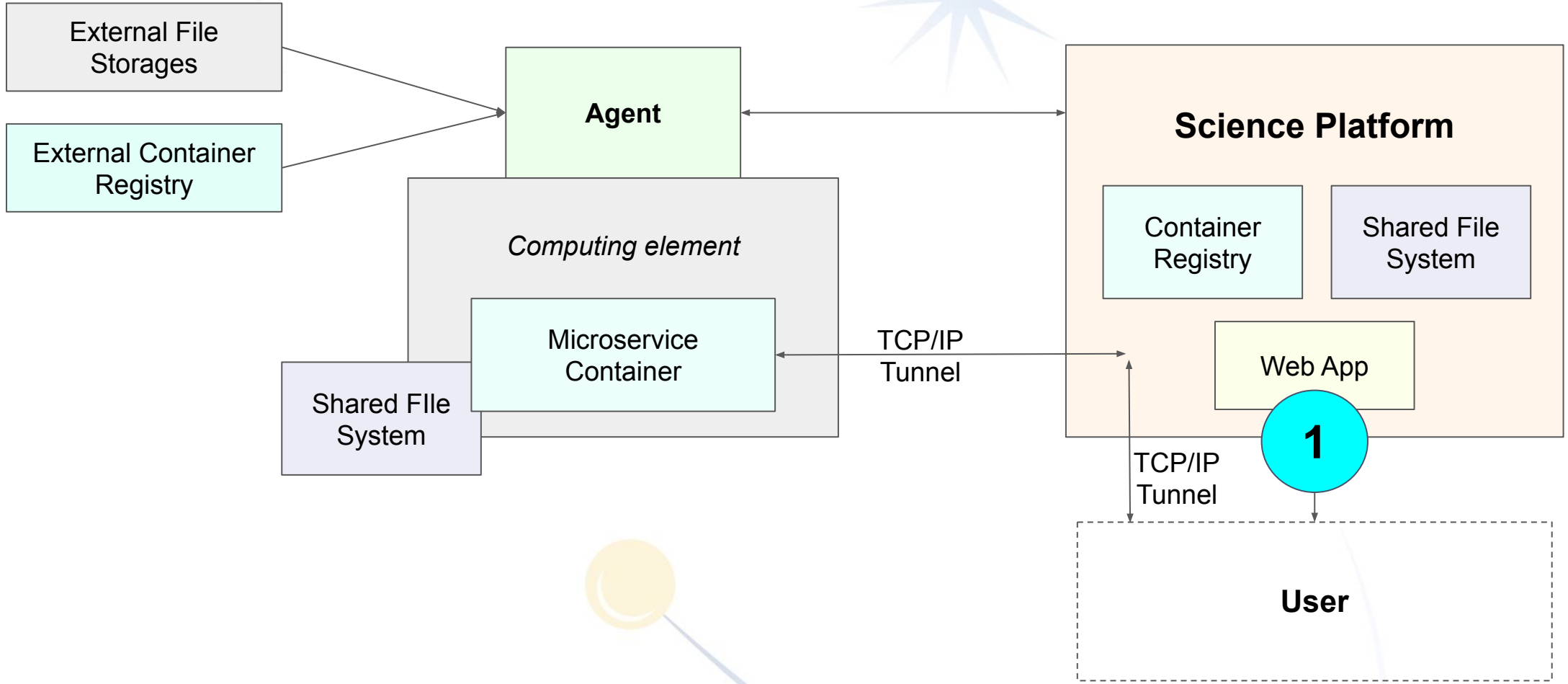
Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta architecture: task orchestration



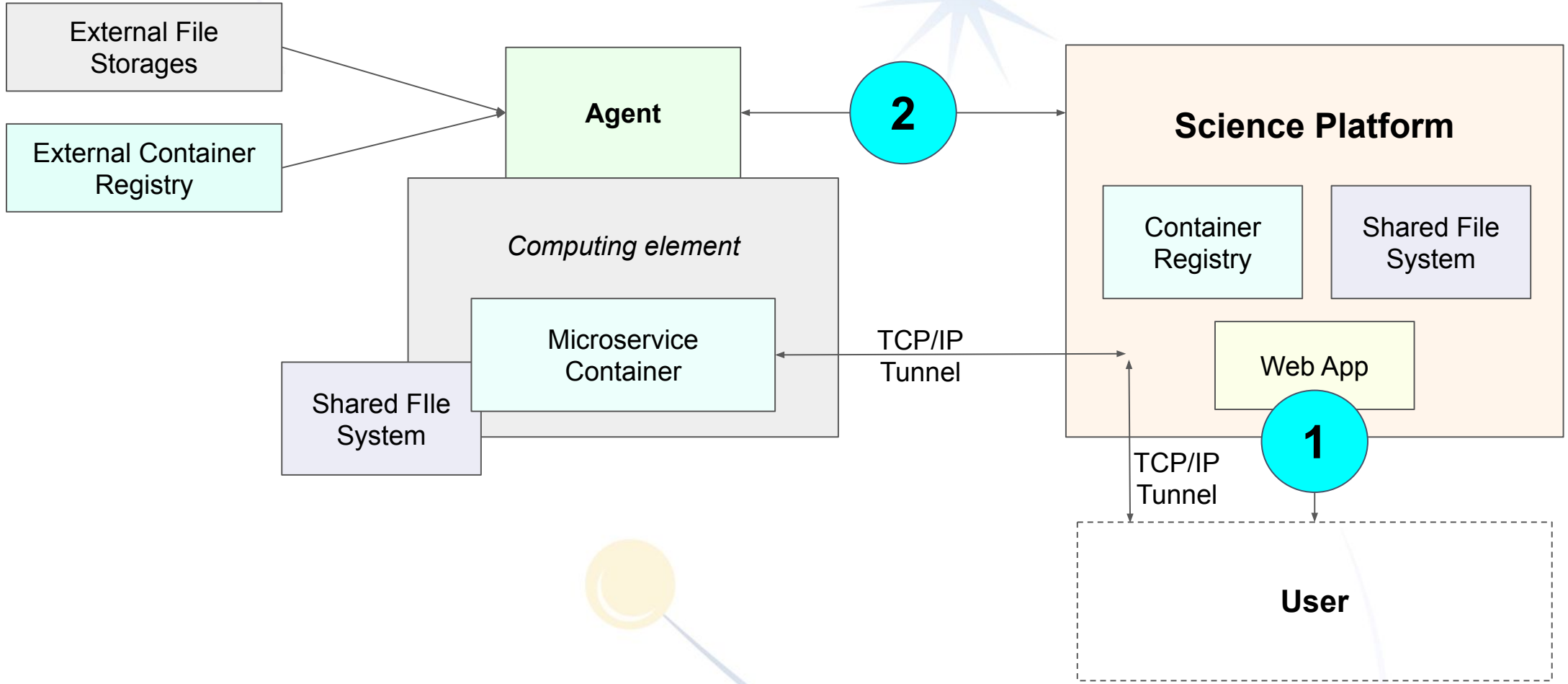
Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta architecture: task orchestration



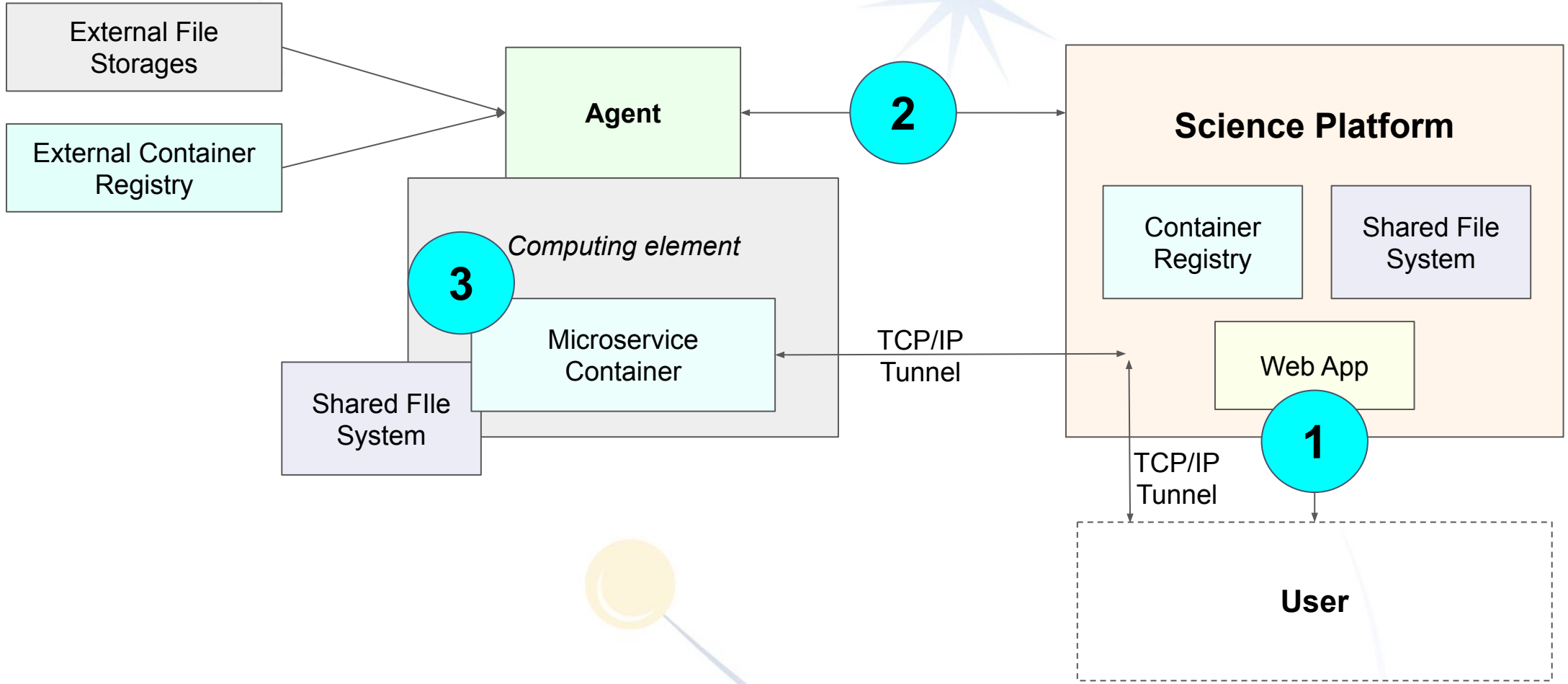
Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta architecture: task orchestration



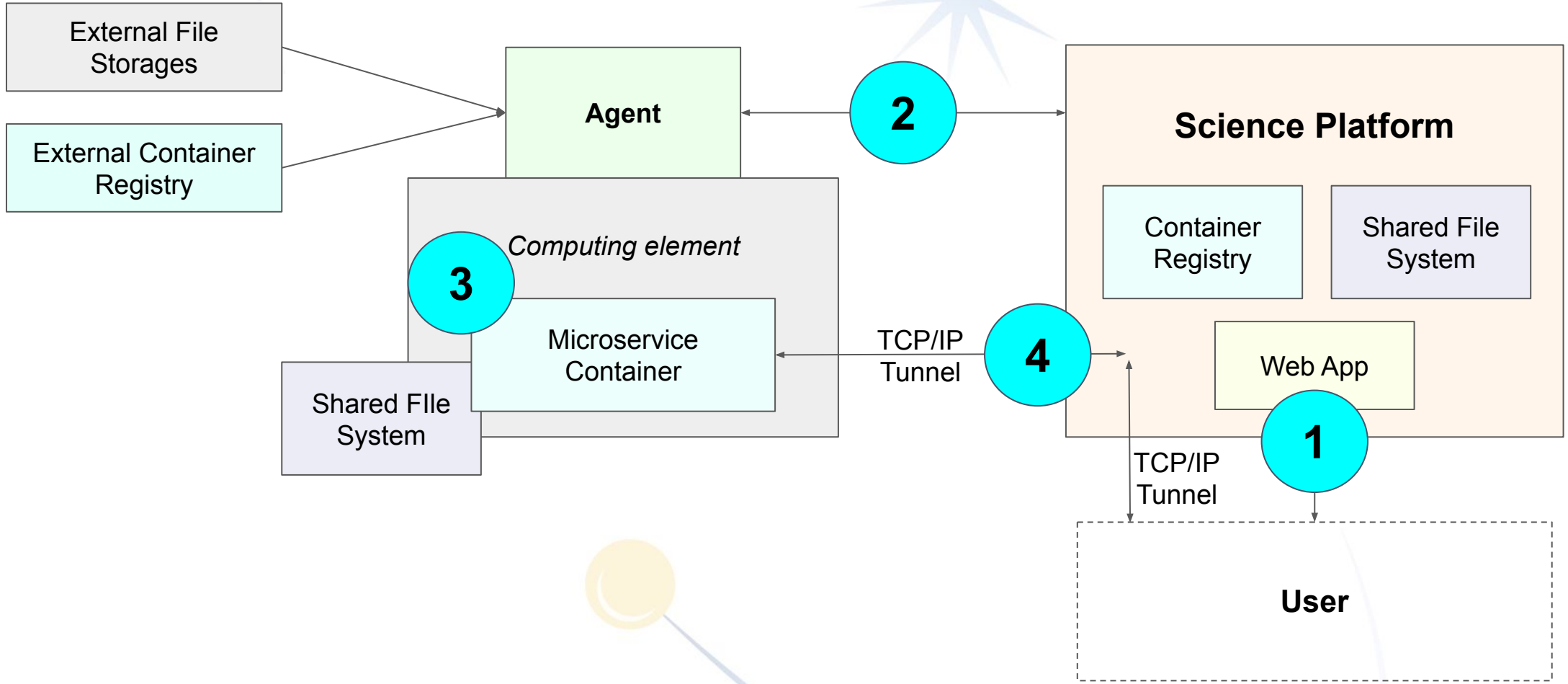
Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta architecture: task orchestration



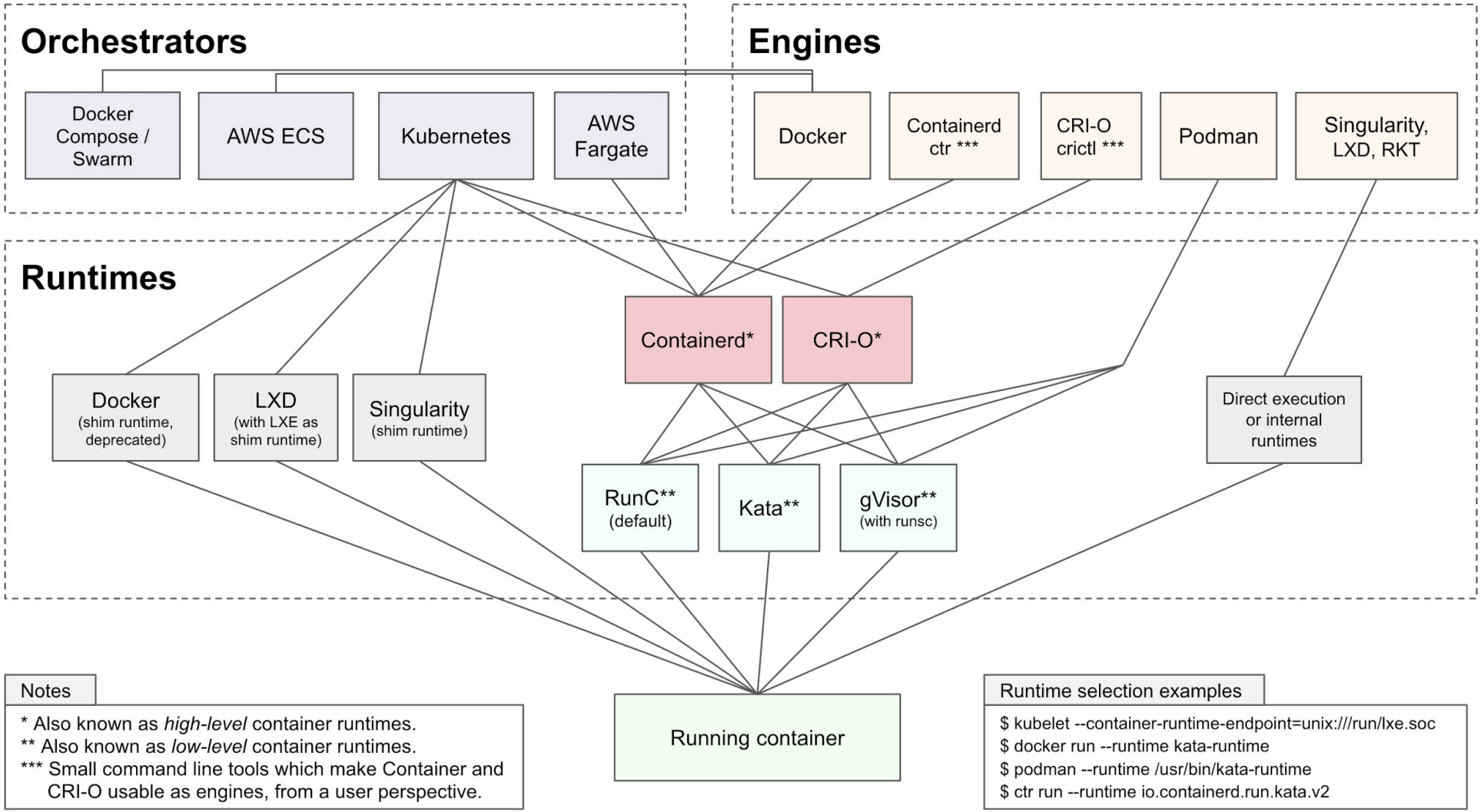
Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta architecture: task orchestration



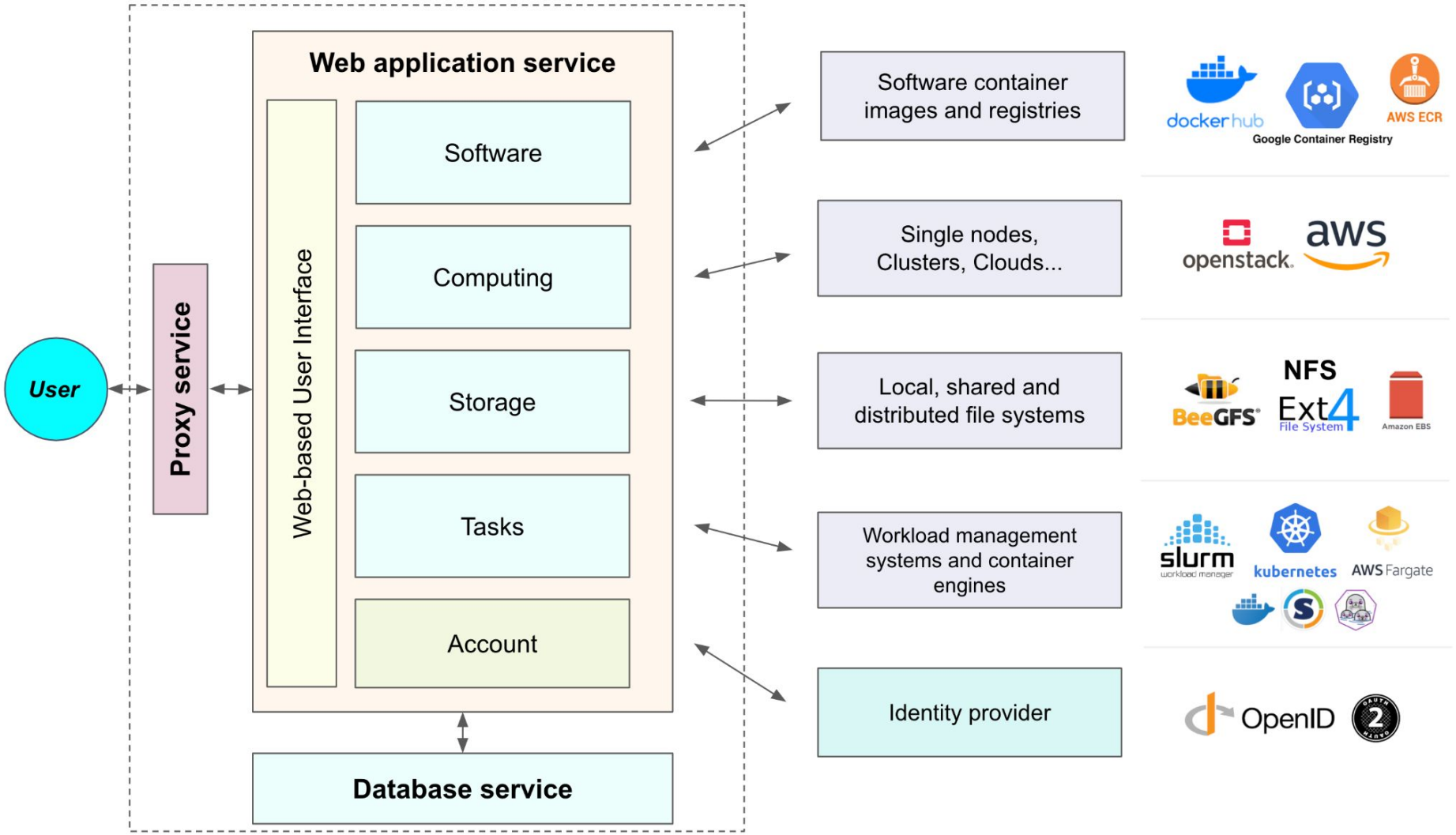
Funded by the European Union's Horizon 2020 - Grant N° 824064

Parenthesis: containerization landscape



Funded by the European Union's Horizon 2020 - Grant N° 824064

Rosetta: an umbrella for *n* technologies



Funded by the European Union's Horizon 2020 - Grant N° 824064

Wrapping up

- **Rosetta** is a **Science Platform** which allows to run:
  - **well established software** *(pre-loaded on the platform)*
  - **custom software** and/or dependencies *(added by the users)*

    ⟶ *both as: Jupyter Notebooks, GUI applications, command line tools etc.*
- Supports **HPC, Cloud** and **data intensive** systems
  - easy to integrate with existing WMS
- **Containers** together with **microservice**-level isolation make it **safe** to let users add their own software and provide a **robust architecture**
- Huge improvements in terms of **reproducibility**

Funded by the European Union's Horizon 2020 - Grant N° 824064

THANKS!

Stefano Alberto Russo - INAF OATS

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.